A Method for Extracting Task-related Information from Social Media based on Structured Domain Knowledge

Submission Type: Full paper

Daniel Link
European Research Center for Information Systems, ERCIS
University of Münster
Münster, Germany
daniel.link@ercis.uni-muenster.de

Flávio E. A. Horita
Department of Computer Systems, ICMC
University of São Paulo
São Carlos/SP, Brazil
horita@icmc.usp.br

João Porto de Albuquerque
Department of Computer Systems, ICMC, University of São Paulo, Brazil
GIScience Chair, Institute Geography, Heidelberg University, Germany
jporto@icmc.usp.br

Bernd Hellingrath
European Research Center for Information Systems, ERCIS
University of Münster
Münster, Germany
bernd.hellingrath@ercis.uni-muenster.de

Shabdiz Ghasemivandhonaryar
University of Münster
Münster, Germany
s_ghas02@uni-muenster.de

Abstract

Social media platforms have come into the focus of research as sources of information about the unfolding situation in disaster contexts. Incorporating information from social media into decision-making is still difficult though. One reason may be that the prevalent approach to data analysis works bottom-up, which has several limitations. In this paper, we adopt a top-down approach by means of a novel keyword-based method for identifying potentially relevant information in social media data based on structured knowledge of activities undertaken in a domain. The application of the method to the context of humanitarian logistics using four social media datasets shows its capability to identify potentially relevant information via reference tasks and to match identified information with decision-makers' activities. In addition, we offer a first set of domain-specific keywords to identify information related to infrastructure and resources in humanitarian logistics.

Keywords (Required)

Disaster management, humanitarian logistics, social media, reference tasks, information categories, infrastructure and resources

Introduction

The number of people vulnerable and affected by disasters has been increasing since the 1970s (Guha-Sapir et al. 2004). Events like the 2010 earthquakes in Haiti and Pakistan, the 2011 Fukushima Daiichi nuclear disaster, and the 2013 typhoon Haiyan in the Philippines caused huge losses in terms of lives and economic damages. In the immediate aftermath of a disaster – during the response phase – action needs to be prompt and targeted, although required information about the situation at hand is least available (Blecken 2010; IFRC 2013). Social media platforms, such as Twitter and Facebook, have come into the focus of research as sources of information about the unfolding situation in disaster contexts (Bakillah et al. 2014; Crooks et al. 2013; Sakaki et al. 2010).

It remains difficult to incorporate information from social media into decision-making (IFRC 2013; Imran et al. 2014; Tapia et al. 2013). One reason may be that the prevalent approach to data analysis works bottom-up, wherein data are analysed first and the target audience's needs are addressed later (Vieweg et al. 2014). While bottom-up analysis helps understanding the decisions of citizens during disasters (Chae et al. 2014; Starbird et al. 2010; Vieweg et al. 2010) and to detect incoming events (Fuchs et al. 2013), it fails to incorporate this information into decision-making processes (Vieweg et al. 2014). Developing a top-down approach seems promising (Vieweg et al. 2014), where relevant questions are identified prior to analysis (Mazurana et al. 2012). Acknowledging that approaches specific to a certain domain often perform better than domain-independent approaches, as is the case with natural language processing applications (Imran et al. 2014), it makes sense to ground a top-down approach in structured domain knowledge.

Therefore, this paper adopts a top-down approach by means of a novel keyword-based method for identifying potentially relevant information in social media data based on structured knowledge of activities undertaken in a domain, and to match identified information with decision-makers' activities. To this end we integrate several fundamental building blocks: reference task modelling and information categories to describe the domain (i.e. to structure the decision space), and keyword-based search as a popular technique to identify relevant content in social media (Imran et al. 2014). We thereby assume that structured domain knowledge is useful for deriving domain-specific keywords, which can in turn be used for targeted social media data analysis and to better link identified information with activities in the targeted domain. The domain of humanitarian logistics represents a suitable case, because of its central role in disaster management and international humanitarian assistance for the successful and fast delivery of aid products and services. Moreover, several reference models and information categories exist that provide structured knowledge of activities in the domain (Link et al. 2015; Widera and Hellingrath 2011).

The remainder of this paper is structured as follows. We first present related work as a backdrop to our study. Next, the developed method is described in a generic way. It is then applied, beginning with an introduction to the chosen domain of humanitarian logistics, including reference tasks and pertinent information categories. On this foundation, sets of keywords are generated and refined. The search results are then controlled for quality and matched with tasks in the application domain. The paper ends with a discussion, including recommendations for future work, and our conclusion.

Related Work

The use of social media platforms for analyzing daily activities and communities' behavior has increased in the last years, from the prediction of stock markets (Bollen et al. 2011) and the creation of brand communities (Muniz, Jr. and O'Guinn 2001) to communities' socialization (e.g. on Facebook). In the context of disasters, some platforms are used for assessment (Guan and Chen 2014), attempting to improve situation awareness (Starbird et al. 2010; Vieweg et al. 2010; Yin et al. 2012; Zielinski et al. 2013) and to predict events (Crooks et al. 2013; Sakaki et al. 2010).

Vieweg et al. (2014) emphasize that a top-down approach to social media analysis requires a sound understanding of decision-makers' information needs and requirements, e.g. regarding the format of provided information and its quality. Interviews of several members of the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) yielded insights of their information needs, which the authors attempt to match with data from social media, esp. Twitter. Although this work provides insights about decision-makers' information needs, the authors do not present any systematic top-down approach.

Most of the existing works in the literature about social media analysis employ keywords merely for querying social media sites, i.e. for data collection (Acar and Muraki 2011; Jadhav et al. 2010; Terpstra et al. 2012). Marcus et al. (2011) make use of keywords for extracting messages and then group and display them in a graphical timeline to support event analysis. Imran et al. (2013) adopt hashtags (keywords preceded by the symbol #) for extracting and estimating the usefulness of messages related to natural disasters. Caragea et al. (2011), Kongthon et al. (2012) and Sakaki et al. (2010) use keywords as a technique for automatically classifying short text messages from the Twitter platform. Hughes et al. (2014) and Dou et al. (2012) use keywords to detect events and trends in data collected with official organizations (e.g. fire departments) during Hurricane Sandy. These works all follow a bottom-up approach, where keywords are defined based on an open-minded exploration of text messages, and the results are then used to some end, like event detection. The work most closely related to this paper – which employs a top-down approach - is *Twitcident*, an information system able to identify real-word incidents and crises using text messages collected on Twitter (Abel et al. 2012b; Abel et al. 2012a). The collected messages are filtered using notifications from official emergency broadcasting services, such as policy departments. Thus, the information that decision-makers receive is strictly filtered, processed, and mainly related to the event. However, Twitcident does not consider the relationship between domain-specific tasks and text messages, which is essential for integrating analysis results with timely and accurate decision-making.

Generic Description of the Method for Extracting Task-related Information

Figure 1 gives a generic overview of the developed method for extracting task-related information, which is described in greater detail in the following.

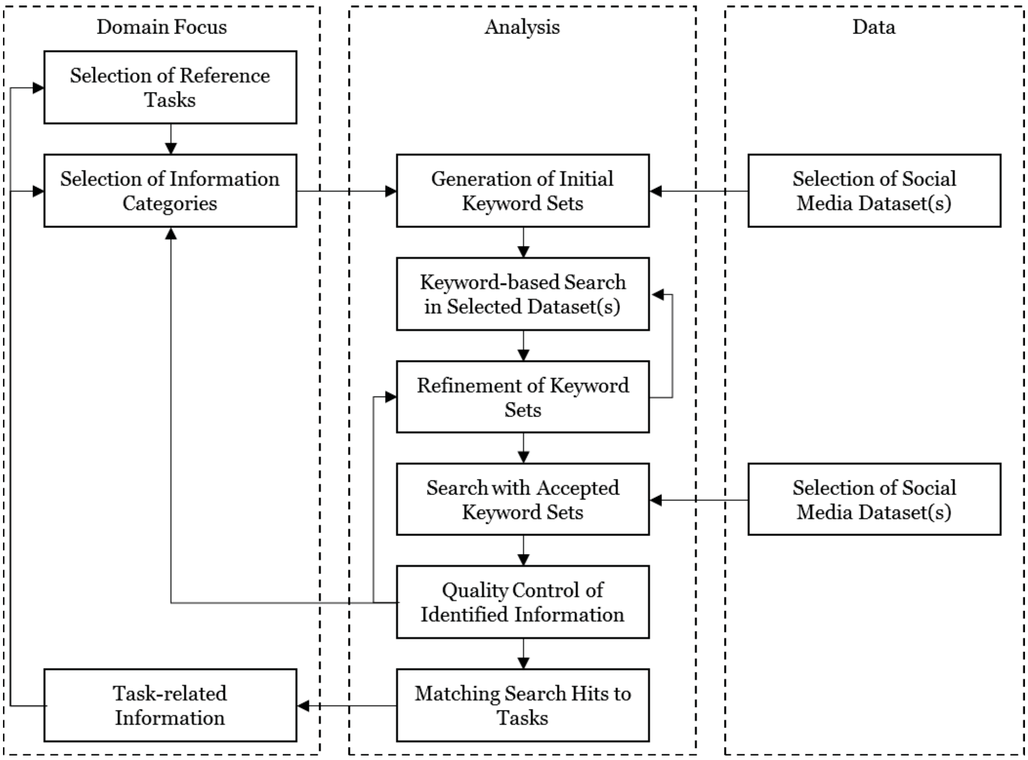

Figure 1. Overview of the Developed Method for Extracting Task-related Information

The process begins by focusing on a domain in terms of selecting reference tasks included in a reference information model (e.g. "Plan Transport Routes") and of information categories pertaining to the selected tasks (e.g. "Road Network"). "*A reference information model is the outcome of the construction of a modeller who declares information about universal elements of a system relevant for application system or organisation developers to solve a specific task at a given time in a given language in such a way that a point of reference for an information system is created*" (Blecken 2010). Information categories refer to the elements of taxonomies or ontologies representing information needs. These can either be drawn from the reference information model if it contains them, or should be constructed on its basis, so that the information categories' proper fit to reference tasks can be ensured.

Additionally, one or more social media datasets are selected for the generation of keyword sets. The derived keyword sets are useful only within the limitations of the selection, which has to be reflected on at this point, e.g. regarding the covered region and type of disaster. For example, a dataset about a disaster in a country with no coastline would unlikely lead to any keywords related to sea transportation.

Next, the selected datasets are analysed in the light of the selected reference tasks and information categories, in order to generate an initial set of keywords per information category. This is initiated by drawing relevant terms directly from the descriptions of reference tasks and information categories, and by including their synonyms. Next, the datasets are manually examined, in order to identify entries that relate to any particular category and to add them to the corresponding keyword set. Ideally, the selected datasets are examined in full. If a dataset is too big for manual examination, however, it is suitable to use a randomized sample. The sample size n can be calculated by using standard equations, such as Equation 1. Therein, N is the population, e is the width of the confidence interval, p is the sampling rate, and Z is the desired confidence.

$$n = \frac{m}{1 + \frac{m-1}{N}} \quad \text{with} \quad m = \frac{z^2 p(1-p)}{e^2}$$

Equation 1. Sample Size (Cochran 1977)

In an iterative process, the generated keyword sets are then used to search the selected datasets to (1) assure that no related entries have been overlooked and (2) to learn which keywords produce false positives. With the gained knowledge, the keyword sets are repeatedly refined and finally accepted.

During search with the accepted keyword sets, each data entry in one or more of the selected datasets that contains an accepted keyword will be placed in one or more categories, depending on the keyword.

For quality control, as is appropriate for an application domain with such high stakes and decision-makers often encountering information overload, human coders manually examine each identified data entry and confirm that it indeed falls into a certain category by coding it accordingly. The reliability of their coding and thus the quality of the categorization can be validated by calculating agreement in terms of inter-coder reliability. Hayes and Krippendorff (2007) analyse existing measures and observe that several of them, like the currently popular Cohen's Kappa that is used by Hughes et al. (2014) and others, have several partly critical disadvantages and are not suitable for the task. Instead, they consider Krippendorff's Alpha to be superior, since it can be used regardless of the number of observers, levels of measurement, sample sizes, and presence or absence of missing data. Most conveniently, they also provide macros for the popular software SPSS and SAS. Although cutting points depend on the context of use, "*to assure that the data under consideration are at least similarly interpretable by two or more scholars (as represented by different coders), it is customary to require $\alpha \geq 0.800$*" (Krippendorff 2004). For this paper, a value of 0.800 shall thus suffice, while the application in a real world context, where lives are at stake, would require a higher value. If quality control fails, the underlying information categories or keyword sets should be adjusted.

Once inter-coder reliability is validated, the search hits can be matched to reference tasks per category, which results in task-related information. The results may be analysed to inform selection of reference tasks and/or information categories and start the process anew.

Application of the Method for Extracting Task-related Information

Domain Focus: Selection of Reference Tasks and Information Categories

Humanitarian logistics, the domain chosen for this work, deals with the process of "*planning, implementing and controlling the efficient, cost effective flow and storage of goods, materials and equipment as well as related information, from point of origin to point of consumption for the purpose of meeting the beneficiary's requirements*" (Blecken 2010). Among the existing reference models that describe the domain's various tasks, roles and variables in a structured way, only Blecken's (2010) model considers all relevant planning horizons as well as operational processes (Widera and Hellingrath 2011). The model categorizes tasks into six functional areas, as seen in Figure 2: assessment, procurement, warehousing, transport, operation support, and reporting. For example, the operational / transport category contains 15 tasks, like *consolidate transport*, *import goods and clear customs*, or *select transport mode*; see the Appendix for exemplary task descriptions.

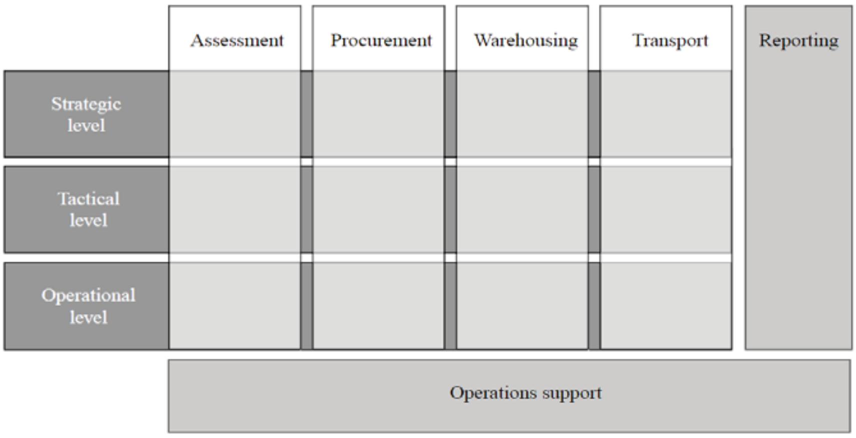

Figure 2. Reference model framework of Blecken (2010)

Humanitarian logisticians need up-to-date and sufficiently accurate information about the infrastructure and resources required for their tasks, like the road network or vehicles. Link et al. (2015) identify and categorize such infrastructure and resources; for example, the main category *Airport* contains sub-categories like *Current Condition of Facilities*, *Runway Characteristics* and *Cost of Airport Usage*.

Table 1 presents an overview of the correspondences between Blecken's functional areas and Link et al.'s infrastructure and resource categories, thereby illustrating which category can provide useful information for a given functional area. For example, information about *Road Transportation* is useful to support a task like S*elect Transport Route*. *Assessment* does not correspond to any category, because its very purpose is to provide required information. *Reporting* neither corresponds to any category, as it depends on data sources internal to the organization. Some categories, like *Country Overview* potentially affect multiple functional areas.

Functional Area (Blecken 2010)	**Infrastructure and Resources Categories (Link et al. 2015)**
Assessment	-
Transport	Airport
	Customs
	General
	Inland waterway transportation
	Railway transportation

Functional Area (Blecken 2010)	Infrastructure and Resources Categories (Link et al. 2015)
	Road transportation
	Seaport
Warehousing	Warehouse
Procurement	Flour
	Local suppliers of relief goods
Reporting	-
Operations Support	Electric power supply
	Fuel supply
	Staff accommodation
	Telecommunications
Multiple areas potentially affected	Buildings
	Additional handling equipment
	Country overview
	Natural environment

Table 1. Functional areas of the reference task model vs. infrastructure and resource categories

Selection of Social Media Datasets for Generation of Keyword Sets

This work uses four social media datasets for generating sets of keywords, as shown in Table 2. These datasets originate from a volunteer & technical community (VTC) and are composed by a distinct number of short text messages, including Twitter messages and SMS text messages. We chose them primarily because they are related to infrastructure and resources in humanitarian logistics. Furthermore, they are heterogeneous in period, location, type of disaster, and number of text messages, which supports the generalization of the developed method.

ID	Event	Location	Timespan (mm/dd/yy)	Messages
H1	Earthquake	Haiti	01/12/10 – 07/05/10	3,594
S	Cyclone Evan	Samoa	12/14/12 – 12/16/12	363
P	Typhoon Pablo	Philippines	Unknown	648
H2	Typhoon Haiyan	Philippines	11/07/13 – 11/08/13	6,027

Table 2. Selected Social Media Datasets

- H1 (Haiti): On January 12th 2010, a magnitude 7 earthquake hit Haiti, affecting more than 5 mn. people, rendering about 1.8 mn. people homeless and causing damages of at least 14 bn. USD (Cavallo et al. 2010; Daniell 2011; Eberhard et al. 2010). Almost all kinds of local infrastructure were affected, including massive damages to roads, bridges and Port-au-Prince's airport and harbour (Eberhard et al. 2010).

- S (Samoa): On December 13th 2012, the category 3 cyclone Evan hit Samoa, causing heavy rainfalls and flooding. The Government of Samoa (2013) reported five people dead, 4,673 people affected and more than 200 mn. USD in damages, of which almost 40 % were related to infrastructure. OCHA reported severe disruptions of power supply on Upolo Island with estimates of up to two months to restore 70 percent of electricity connections. All ports were closed. 50% of roads on Upolo Island have sustained damages, with 60% being passable at the time of the report (OCHA 2012a).
- P (Pablo): On December 4th 2012, the category 5 typhoon Pablo hit the Philippines. The Government of the Philippines (2012) reported 6.2 mn. people affected, about 1,000 dead, 2,622 injured and 841 missing. Pablo caused 156 bn. USD damages to local infrastructure. OCHA stated the need to create logistics centres with proper information technologies and warehouses, as well as the requirement to capture, store and transport relief goods to affected areas (OCHA 2012b, 2013a).
- H2 (Haiyan): On November 8th 2013, the category 5 typhoon Haiyan hit the Philippines, having 14 mn. people affected, 6,200 dead, 26,000 heavily injured, 1,613 missing, and 3.5 mn. Displaced. OCHA (2013b) further reported logistics challenges relating to low truck capacities, lack of fuel and too few warehouses.

Generation of Initial Keyword Sets

The initial sets of keywords are founded on descriptions of reference tasks and information categories. Details are added through manually examining text messages in the datasets[1] and assigning them to infrastructure and resource categories. Table 3 shows exemplary keyword sets for the information category *Airport* in the functional area *Transport* and for *Fuel Supply* in *Operations Support*; see the appendix for all generated keyword sets and an explanation of their syntax.

Functional Area	**Information Category**	**Keyword Set**
Transport	Airport	Air Cargo, aircraft, airport, airplane, flight (schedule), helicopter, landing, sites, and planes.
Operations Support	Fuel Supply	Crude oil, fuel (supply), furnish, gas (station), gasoline, kerosene, naphtha, petrol, petroleum, unfuelled.

Table 3. Exemplary keyword sets for the information categories *Airport* and *Road Transportation*

Manual examination did not reveal any text messages relating to the following categories: additional handling equipment, customs, flour, local suppliers of relief goods, staff accommodation. However, these categories are still regarded in the next step, in order to reduce the probability of text messages skipping manual examination.

[1] For datasets with less than 1,000 text messages (S, P) we examined each text message. For bigger datasets (H1, H2) we used a randomized sample, in accordance with the generic method and using Equation 1 to calculate sample size. As the sampling rate is unknown, we assume $p=0.5$, i.e. half of the information is guessed to be (ir)relevant. Our desired confidence is 95%, i.e. $Z=1.96$. This results in the sample sizes 823 (H1) and 906 (H2). Examination of dataset H2 showed a relatively low share of useful information, so we enlarged the sample size from 906 to 1,300, giving results of a similar quality as the other datasets.

Keyword-based Search in Selected Datasets

During keyword-based search, the dictionary entries are used as keywords to search the complete datasets, in order to identify more potentially relevant text messages. Newly identified text messages are manually examined and the process iterates from there until no more relevant text messages can be identified.

Using the constructed keyword sets to search the complete datasets, we found 3,226 hits (30.34% of all text messages), of which eventually 842 (7.92% of all text messages) were considered relevant for keyword generation. The following two messages are examples for potentially relevant information.

- "*PaP Airport Update, more detailed info. Boeing 757s can land, the runway is 141-feet wide.*" (H1)
- "*Vessels cannot deliver fuel to the port's tank farm due to damage to the fuel-delivery pier, and the report noted an estimated two-day supply remains in bulk storage before fuel becomes scarce.*" (H1)

Refinement of Keyword Sets

To refine the constructed keyword sets, their occurrence and distribution over categories is analyzed, which reveals potential relations among individual keywords and information categories as well as in-between information categories.

We found categories relating to the transport function to overlap considerably, so we grouped the overlapping terms into the separate category *general transportation*.

The datasets contain text messages with contextual information that does not easily fit into Link et al.'s (2015) information categories but does have potential impact on operations; e.g., rubble from buildings potentially blocks roads, and extreme weather in a certain region can impact staff safety. To maintain a clear view on such information without contaminating the original category system, they are grouped in two additional categories: *buildings* and *natural environment*.

Most messages that the final keyword sets lead to fall into the categories *buildings* and *natural environment*. From the initial categories, *electric power supply*, the group of transport-related categories, and *telecommunications*. Most notably, several messages could be identified that relate to *local suppliers of relief goods*, which was regarded during generation of keyword sets although no related messages could be identified at that stage.

Some keywords, like *house*, *home*, *airport*, *fuel* or *road* generate many irrelevant hits, and individual keywords consistently lead to more ambiguous information. For instance, *road* often refers to road damages but is also part of directions, as in "on the road to Jacmel", or used metaphorically. In contrast, keyword combinations have a higher chance to lead to relevant findings, such as *road blocked*. For example:

- "*We are near Palmiste Tavin, Leogane on the road to Jacmel. We need help we don't have food or water*" (H1)
- "*Toai - Cross Island Rd - 5 families stuck - road blocked by trees*" (S)

Quality Control of Identified Information

Three people followed identical instructions to independently code messages in all four datasets. The codings lie on a nominal scale. When a text message was coded with multiple categories, it was considered separately for each category. The computed reliability scores (see Table 4) all satisfy Krippendorff's (2004) requirement of $\alpha \geq 0.800$ for scholarly discussion.

ID	Alpha
H1	0.9019
S	0.8293
P	0.8269
H2	0.9272

Table 4. Inter-coder reliability

Matching Search Hits to Tasks

To establish a relation of categorized text messages to tasks, it is necessary to match them via information categories with functional areas and subsequently identify potentially affected tasks within these areas. The following examples illustrate the process; see the Appendix for selected reference task descriptions.

- "*PaP Airport Update, more detailed info. Boeing 757s can land, the runway is 141-feet wide.*" (H1)
  - Information category: airport
  - Functional area: transport
  - Potentially affected tasks: (1) plan transport mode, (2) select transport mode, (3) plan transport routes, (4) select transport route
- "*Vessels cannot deliver fuel to the port's tank farm due to damage to the fuel-delivery pier, and the report noted an estimated two-day supply remains in bulk storage before fuel becomes scarce.*" (H1)
  - Information category: fuel supply
  - Functional area: operations support
  - Potentially affected task: prioritize and allocate operations support resources

Discussion

We designed the method presented in this paper as a top-down approach to analysis of social media data from disaster contexts, in order to identify relevant information in text messages and match identified messages with tasks of decision-makers. A general limitation of using social media data to inform decision-makers in a particular domain lies in the availability of needed information to the social media users who post messages, and in their interests. For example, it can be safely assumed that most users are hardly interested to talk about customs or staff accommodation, and we did indeed not find any such information in the selected datasets. In the future, it may become possible to rapidly process social media data into a relevant picture of the operational environment, which comes to responding organizations in a standardized format that makes sense to them (Tapia et al. 2013). To achieve this vision of a relevant picture of the operational environment, we see the need to integrate more data sources into analysis than just text messages. This raises questions regarding the integration of heterogeneous sources, especially of hard and soft sensor data, which is still difficult (Hall and Jordan 2010; Horita et al. 2015).

The datasets selected for generating keywords focus on certain kinds of events, i.e. they ignore famines, acts of terrorism, civil unrest, etc. This and the relatively big size of datasets H1 and H2 have likely introduced a bias to the keyword sets. Despite not affecting the proof-of-concept undertaken in this work, it means that the biases resulting from the datasets should be carefully reflected on before further application, as mentioned already in the section describing the general search method. Future work should consider further datasets, such as the ones provided by public repositories of crisis-related data; see CrisisLex.org and CrisisNET.

Furthermore, text messages typically are highly contextual, and the employed language is often highly ambiguous. Individual keywords tend to lead to more ambiguous findings, while certain keyword combinations exhibit a higher hit ratio. It seems promising to employ an approach like Lexical Link Analysis (Zhao et al. 2013) to enable a more in-depth study of the relation between keywords, eventually better informing monitoring techniques like Twitcident's semantic analysis (Abel et al. 2012b; Abel et al. 2012a). Other methods that can support semantic analysis include techniques from natural language processing. Regarding the variety of expressions used in social media, Imran et al. (2014) point out that it is not sufficient to only look for conceptual keywords, like *airport*. This is because mentions of particular instances of concepts, e.g. *JFK* for John F. Kennedy International Airport, are not detected. They suggest named entity linking as one widely used semantic technology, which looks up particular instances and their classes (e.g. on Wikipedia) and links them to an ontology. Future efforts can be made to combine our search method and named entity linking via Link et al.'s (2015) information categories. This promises improvements in categorizing information, which increases the volume of information matched to reference tasks.

Applying our method is a labour-intensive process that includes much manual work, which calls for automation. There are two fundamentally different approaches to take towards (semi-)automated interactive analytics. Like Endert et al. (2014), we believe that human-in-the-loop thinking in interactive analytics inevitably leads to usability problems. This is because end users "*are presented with results out of context, without understanding their meaning or relevance, and interactive controls are algorithm specific and difficult to understand*" (Endert et al. 2014). Instead, a human-is-the-loop kind of approach to interactive analytics promises better results but requires an understanding of human work processes, which can lead to the design of better work process and fitting algorithms – possibly from the ground up. Reference tasks provide an understanding of human work processes that is useful for initial method development. To further improve the developed method and to achieve targeted real-time analysis, it will be necessary to work with decision-makers (e.g. logisticians) and providers of analysis services (e.g. volunteer & technical communities/VTCs) to design a better process for targeted, real-time analysis. This may include the design of sectorial intelligence services that anticipate information needs in the field[2], handle requests for information (i.e. decision-makers pull information), provide more concrete guidance to VTCs, and act as filters for push-based information distribution to the field. Improved work processes can then be the basis for automation, e.g. by utilizing supervised machine learning. One approach is to use the keyword-based search to identify data entries that can serve as a gold standard, training the machine learning system. The system would, in turn, suggest further data entries that match the identified categories. In this way, human and machine would work together to improve the generated keyword sets.

The high stakes in disaster management and humanitarian assistance demand proper testing of any solution, before it can be responsibly introduced in real world situations. Serious games, whose design can be based on reference tasks (Link et al. 2014), present an interesting opportunity to evaluate the developed process and supporting information system in an intermediary step, before deploying in full-sized exercises and eventually real disaster contexts.

To allow the generalization of this work, applications to further domains are necessary. This is due to the multitude of critical factors involved in a real disaster. As another application domain, we suggest the area of monitoring of environmental variables and early warning, up-to-date information is essential for effective disaster preparedness and the timely triggering of rapid assessment and response.

---

[2] Regarding information managers anticipating information needs of field logisticians, see the chapter on information management in Tomasini and Van Wassenhove (2009).

Conclusion

This paper presents a method for performing keyword-based search of task-related information in social media data. Our foundations are structured domain knowledge in the form of reference tasks and information categories. To the best of our knowledge, the developed method represents the first systematic top-down approach to social media analysis in disaster management, wherein structured domain knowledge leads to relevant findings and matches identified information with activities undertaken in the domain. This stands in contrast to the bottom-up construction of categories through data exploration and the subsequent struggle to tap into important workflows in the targeted domain. The biggest benefit of this work thus lies in an easy matching of identified information via reference tasks to decision-makers' activities, thereby reducing the gap between social media analysis and decision-making. While much work remains to be done to finally close this gap, the application of the developed method to the domain of humanitarian logistics proves the concept by providing task-related information from social media datasets, and offers a first set of domain-specific keyword sets for the social media analysis targeting this domain. Future work may focus on the integration of heterogeneous sources, consider public crisis data repositories, improve accuracy through better semantic analysis, and take a human-is-the-loop approach to the design of better work processes and eventually to automation.

Acknowledgements

FEAH and JPA would like to express thanks for the financial support provided by the FAPESP-IVA project "Assessment of Impacts and Vulnerability to Climate Change in Brazil and Strategies for Adaptation Options" (Grant no. 2008/58161-1) and CAPES (Grant Edital Pró-alertas no. 24/2014). FEAH is grateful for the financial support from CNPq (Grant no. 202453/2014-6). JPA is grateful for the financial support from FAPESP (Grant no. 2012/18675-1), CAPES (Grant no. 12065-13-7) and Heidelberg University (Excellence Initiative fund: 2300054, assignment: 7812617).

Appendix

Exemplary Reference Task Descriptions

- Transport – **Select Transport Mode**: "*A suitable transport mode must be selected if several modes of transport are available. Within a multi-echelon humanitarian supply chain, the last stage usually consists of road transport, although other modes of transport such as small boats or air drops are conceivable. When selecting a transport mode, several factors, such as cost, lead time, speed, flexibility, and robustness need to be taken into account. Hazardous materials or material requiring special handling, restrict suitable transport modes.*" (Blecken 2010)
- Operations Support (OS) – **Prioritize and Allocate OS Resources**: "*This task precedes all tasks occupied with the mobilisation of resources. OS resources are scarce and need to be prioritised in case operations are being conducted in various regions. The available OS resources need to be (pre-) allocated to these regions and, thus, set boundaries within which ad hoc mobilisation can take place. In order to integrate its findings, prioritisation may need to be re-run after the initial assessment has been concluded.*" (Blecken 2010)

Sets of Keywords Relating to Infrastructure and Resources Relevant to Humanitarian Logistics

Table 5 presents the entire list of generated keyword sets, including their relation to functional divisions and the information categories used for this study or resulting from it, respectively. The keyword sets are described following a certain syntax. Terms without quotation marks are used as exact search terms. Parentheses, as in "member(s)", denote both "member" and "members". Terms separated by slashes indicate possible word combinations, like "building (broken / collapsed)" denotes the following set of terms considered: "building", "building broken", and "building collapsed". Alternative spellings, e.g. "harbor" and "harbour", were considered in the analysis but are omitted here for the sake of concise display and readability.

Functional Areas	Information Categories	Constructed Dictionaries
Transport	Airport	Air Cargo, aircraft, airport, airplane, flight (schedule), helicopter, landing, sites, and planes.
	Customs	Border (control / patrol / protection), contraband, custom duties, customs, frontier guard, import duty, pay duties, regulation, smuggling, tariff.
	General transportation	Bridge (collapsed / damaged), cannot get through, fallen power pole, fallen tree, gating, obstruction.
	Inland waterway transportation	(Ferry / container / inland) harbour, (domestic / ferry / inland) port, river, ship, shipyard, (inland) waterway, inland (water vessel / waterway craft), (river) barge, (inland / fresh / continental) water.
	Road transportation	Blocking road, buggy, bus, cargo, downtown, fallen on road, fell on road, jeep, lorry, on the street, pickup truck, road (blocked / damage(d) / obstruction), roadster, station wagon, truck, vehicle.
	Railway transportation	Main track, platform, railroad (depot), rails, rail track, railway (station), train (station), trench station.
	Seaport	Abyss, coast, container harbour, deep, dock (yard), domestic port, ferry (harbour / port), harbour, inland harbour, inland port, marine, navy, ocean, river, sea harbour, sea transport, ship (yard), streamlet, vessel, waterway.
Warehousing	Warehousing	Warehouse, stockroom, storehouse, store, depot, depository, repository, stock, storage, inventory, reservoir, stocking, shed, stockpile, entrepot, goods depot, storeroom, storage capacity, reserve, savings.
Procurement	Flour	Cereal, corn, crop, farina, flour, grain, meal, mill, milling, ore.
	Local suppliers of relief goods	External provider / supplier / vendor, humanitarian (aid / good), local supplier, provider, relief good, vendor.
Operations Support	Electric power supply	Blackout, broken power line, electric pole(s), electric power/wire, electricity, (need of) electricity, no power, power off / outage / plant / poles / supply, power line(s), supply, wire.
	Fuel supply	Crude oil, fuel (supply), furnish, gas (station), gasoline, kerosene, naphtha, petrol, petroleum, unfuelled.

Functional Areas	Information Categories	Constructed Dictionaries
	Staff accommodation	Accommodation, aid organization, collaborator, lodgement, lodging, member(s) of staff, occupancy, sanitary, settlement.
	Telecommunications	Communication (issue / line / impact), globe, phone, smart, telecom, telecommunication, telephone (mast / pole).
Multiple divisions potentially affected	Additional handling equipment	Appropriation, component, conveyor, elevator, fork lift, forklift, haul means, haulage means.
	Buildings	Building (broken / collapse / collapsed / crumbled / crushed / damage(d) / destroyed / on fire), collapsed structure, construction, factory, (fell / fallen) on house, garden, home (is) (collapsed / crumbled / destroyed / on fire), house (affected / broken / collapsed / crumbled / crushed / damage / damaged / destroyed / toppled), house(s) (destroyed / on fire / affected), park, roof, school (building), structure fallen, teardown, trapped in, university, wreckage, yard.
	Country overview	Armed, arms, army, clash, conflict, flee, overpopulation, political, politics, population, prosecution.
	Natural environment	Area under super-typhoon, earthquake in, epicentre, epicentrum, expected to make landfall, eye of Typhoon, impact expected, landfall (estimated), located (at), Location of Typhoon, made landfall, path (moved), Typhoon at, was located, intensity rating, intensity scale, "km/h", kmh, knots, kt, kts, maximum intensity, maximum sustained winds, mb, mbar, mph, pressure, Richter scale, speed, sustained winds, wind speed, aftershock, caution, conjecture, cyclone warning, estimates, expectation, forecast, prediction, prognosis, warning signal, (yellow) rainfall warning.

Table 5. Information categories with their information categories and corresponding functional divisions

REFERENCES

Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., and Tao, K. 2012a. "Twitcident: Fighting Fire with Information from Social Web Streams," in *Proceedings of the 21st annual Conference on World Wide Web Companion,* Lyon, France. 16-20.04.2012, pp. 305–308.

Abel, F., Hauff, C., Houben, G.-J., Tao, K., and Stronkman, R. 2012b. "Semantics + Filtering + Search = Twitcident: Exploring Information in Social Web Streams," in *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, Milwaukee, WI, pp. 285–294.

Acar, A., and Muraki, Y. 2011. "Twitter for crisis communication: lessons learned from Japan's tsunami disaster," *International Journal of Web Based Communities* (7:3), p. 392.

Bakillah, M., Li, R.-Y., and Liang, S. H. 2014. "Geo-located community detection in Twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon Haiyan," *International Journal of Geographical Information Science* , pp. 1–22.

Blecken, A. 2010. *Humanitarian Logistics*: *Modelling supply chain processes of humanitarian organisations*, Bern: Haupt.

Bollen, J., Mao, H., and Zeng, X. 2011. "Twitter mood predicts the stock market," *Journal of Computational Science* (2:1), pp. 1–8.

Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H.-W., Mitra, P., Wu, D., Tapia, A. H., Giles, L., Jansen, B. J., and Yen, J. 2011. "Classifying Text Messages for the Haiti Earthquake," in *Proceedings of the 8th International ISCRAM Conference,* Lisbon, Portugal.

Cavallo, E., Powell, A., and Becerra, O. 2010. "Estimating the Direct Economic Damages of the Earthquake in Haiti," *The Economic Journal* (120:546), pp. F298–F312.

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., and Ebert, D. S. 2014. "Public behavior response analysis in disaster events utilizing visual analytics of microblog data," *Computers & Graphics* (38), pp. 51–60.

Cochran, W. G. 1977. *Sampling techniques*, New York: Wiley.

Crooks, A., Croitoru, A., Stefanidis, A., and Radzikowski, J. 2013. "#Earthquake: Twitter as a Distributed Sensor System," *Transactions in GIS* (17:1), pp. 124–147.

Daniell, J. 2011. "CATDAT Damaging Earthquakes Database: 2010 - The Year in Review," *CEDIM Earthquake Loss Estimation Series* Research Report 2011-01, Centre for Disaster Management and Risk Reduction Technology (CEDIM), Karlsruhe.

Dou, W., Wang, X., Skau, D., Ribarsky, W., and Zhou, M. X. 2012. "LeadLine: Interactive visual analysis of text data through event identification and exploration," in *Proceedings of the 2012 IEEE Conference on Visual Analytics Science & Technology*, M. Ward and G. Santucci (eds.), Seattle, WA. 14.-19.10.2012, Piscataway, N.J.: IEEE, pp. 93–102.

Eberhard, M. O., Baldridge, S., Marshall, J., Mooney, W., and Rix, G. J. 2010. "The Mw 7.0 Haiti earthquake of January 12, 2010:USGS/EERI advance reconnaissance team report: Open-File Report 2010-1048," U.S. Department of the Interior; U.S. Geological Survey.

Endert, A., Hossain, M. S., Ramakrishnan, N., North, C., Fiaux, P., and Andrews, C. 2014. "The human is the loop: new directions for visual analytics," *Journal of Intelligent Information Systems* (43:3), pp. 411–435.

Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S., and Stange, H. 2013. "Tracing the German centennial flood in the stream of tweets," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, D. Pfoser and A. Voisard (eds.), Orlando, FL. 05.11.2013, pp. 31–38.

Government of Samoa 2013. "Post-disaster Needs Assessment Cyclone Evan 2012," Ministry of Finance, Apia, Samoa.

Government of the Philippines 2012. *Situation report on the effects of typhoon Pablo (NDRRMC), as of December 20, 2012, 7 a.m.* http://www.gov.ph/2012/12/20/situation-report-on-the-effects-of-typhoon-pablo-ndrrmc-as-of-december-20-2012-7-a-m/. Accessed 19 November 2014.

Guan, X., and Chen, C. 2014. "Using social media data to understand and assess disasters," *Natural Hazards* (74:2), pp. 837–850.

Guha-Sapir, D., Hargitt, D., Hoyois, P., Below, R., and Brechet, D. 2004. *Thirty years of natural disasters 1974-2003: The numbers*, Brussels, Belgium: University of Louvain Presses.

Hall, D. L., and Jordan, J. M. 2010. *Human-centered information fusion*, Boston: Artech House.

Hayes, A. F., and Krippendorff, K. 2007. "Answering the Call for a Standard Reliability Measure for Coding Data," *Communication Methods and Measures* (1:1), pp. 77–89.

Horita, F. E., Albuquerque, J. P. de, Degrossi, L. C., Mendiondo, E. M., and Ueyama, J. 2015. "Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks," *Computers & Geosciences* (80), pp. 84–94.

Hughes, A. L., St. Denis, L. A. A., Palen, L., and Anderson, K. M. 2014. "Online public communications by police & fire services during the 2012 Hurricane Sandy," in *Proceedings of the 32nd annual ACM Conference on Human Factors in Computing Systems,* Toronto, Canada. 26.04.-01.05.2014, pp. 1505–1514.

IFRC 2013. "World Disasters Report 2013: Focus on technology and the future of humanitarian action," International Federation of Red Cross and Red Crescent Societies, P. Vinck (ed.), Geneva.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S. 2014. "Processing Social Media Messages in Mass Emergency: A Survey," *ArXiv e-prints* (1407.7071).

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P. 2013. "Practical Extraction of Disaster-Relevant Information from Social Media," in *Proceedings of the 22nd international conference on*

World Wide Web companion, D. Schwabe (ed.), International World Wide Web Conferences Steering Committee, pp. 1021–1024.

Jadhav, A. S., Purohit, H., Kapanipathi, P., Anantharam, P., Ranabahu, A. H., Nguyen, V., Mendes, P. N., Smith, A. G., Cooney, M., and Sheth, A. P. 2010. "Twitris 2.0: Semantically Empowered System for Understanding Perceptions From Social Data," The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis).

Kongthon, A., Haruechaiyasak, C., Pailai, J., and Kongyoung, S. 2012. "The role of Twitter during a natural disaster: Case study of 2011 Thai Flood," in *2012 Portland International Conference on Management of Engineering and Technology*, Piscataway: IEEE, pp. 2227–2232.

Krippendorff, K. 2004. "Reliability in Content Analysis," *Human Communication Research* (30:3), pp. 411–433.

Link, D., Hellingrath, B., and Bültemann, C. 2015. "Information Categories for Infrastructure and Logistic Resource Assessments in Humanitarian Logistics: Exploring Practical Sources with Inductive Category Development.," in *Logistics Management*: *Products, Actors, Technology - Proceedings of the German Academic Association for Business Research, Bremen, 2013*, J. Dethloff, H.-D. Haasis, H. Kopfer, H. Kotzab and J. Schönberger (eds.), Berlin, Heidelberg: Springer, pp. 445–453.

Link, D., Meesters, K., Hellingrath, B., and Van de Walle, Bartel 2014. "Reference Task-based Design of Crisis Management Games," in *Proceedings of the 11th International ISCRAM Conference*, S. R. Hiltz, M. S. Pfaff, L. Plotnick and P. C. Shih (eds.), State College, PA. 18.-21.05.2014, pp. 592–596.

Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., and Miller, R. C. 2011. "Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration," in *Proceedings of the 2011 annual Conference on Human Factors in Computing Systems*, D. S. Tan, G. Fitzpatrick, C. Gutwin, B. Begole and W. A. Kellogg (eds.), Vancouver, BC. 07.-12.05.2011, New York: ACM Press, p. 227.

Mazurana, D., Benelli, P., Gupta, H., and Walker, P. 2012. "Sex and Age Matter: Improving Humanitarian Response in Emergencies," Feinstein International Center, Tufts University.

Muniz, A. M., Jr., and O'Guinn, T. C. 2001. "Brand Community," *Journal of Consumer Research* (27:4), pp. 412–432.

OCHA 2012a. "Pacific: Tropical Cyclone Evan: Situation Report No. 2 (as of 15 Dec 2012)," United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

OCHA 2012b. "Philippines: Typhoon Bopha: Situation Report No. 7 (as of 14 December 2012)," United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

OCHA 2013a. "Philippines: Typhoon Bopha Situation Report No.19 (as of 12 February 2013)," United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

OCHA 2013b. "Philippines: Typhoon Haiyan: Situation Report No. 18 (as of 27 November 2013)," United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

Sakaki, T., Okazaki, M., and Matsuo, Y. 2010. "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," in *Proceedings of the 19th international conference on World wide web*, New York: Association for Computing Machinery (ACM), pp. 851–860.

Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. 2010. "Chatter on the red: what hazards threat reveals about the social life of microblogged information," in *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, K. Inkpen (ed.), New York, NY: ACM Press, p. 241.

Tapia, A. H., Moore, K. A., and Johnson, N. J. 2013. "Beyond the Trustworthy Tweet: A Deeper Understanding of Microblogged Data Use by Disaster Response and Humanitarian Relief Organizations," in *Proceedings of the 10th International ISCRAM Conference*, T. Comes, F. Friedrich, S. Fortier, J. Geldermann and T. Müller (eds.), Baden-Baden, pp. 770–779.

Terpstra, T., Vries, A. de, Stronkman, R., and Paradies, G. L. 2012. "Towards a realtime Twitter analysis during crises for operational crisis management," in *Proceedings of the 9th International ISCRAM Conference*, L. Rothkrantz, J. Ristvej and Z. Franco (eds.), Vancouver, Canada.

Tomasini, R., and Van Wassenhove, L. N. 2009. *Humanitarian Logistics*, UK: Palgrave Macmillan.

Vieweg, S., Castillo, C., and Imran, M. 2014. "Integrating Social Media Communications into the Rapid Assessment of Sudden Onset Disasters," in *Proceedings of the 6th International Conference on Social informatics*, L. M. Aiello and D. McFarland (eds.), Barcelona, Spain. 11.-13.11.2014, pp. 444–461.

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. 2010. "Microblogging during two natural hazards events: What Twitter May Contribute to Situational Awareness," in *Proceedings of the 28th international conference on Human factors in computing systems*, E. Mynatt (ed.), Atlanta, Georgia, New York, NY: ACM Press, pp. 1079–1088.

Widera, A., and Hellingrath, B. 2011. "Improving Humanitarian Logistics - Towards a Tool-based Modeling Approach," in *Logistikmanagement*: *Herausforderungen, Chancen und Lösungen*, E. Sucky, B. Asdecker, A. Dobhan, S. Haas and J. Wiese (eds.), Bamberg: Univ. of Bamberg Press, pp. 273–295.

Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R. 2012. "Using Social Media to Enhance Emergency Situation Awareness," *IEEE Intelligent Systems* (27:6), pp. 52–59.

Zhao, Y., MacKinnon, D., and Brutzman, D. 2013. *Lexical Link Analysis (LLA) and MMOWGLI*. https://portal.mmowgli.nps.edu/game-wiki/-/wiki/PlayerResources/Lexical+Link+Analysis/. Accessed 13 November 2014.

Zielinski, A., Tokarchuk, L., Middleton, S., and Chaves, F. 2013. "A Framework for Achieving Situational Awareness during Crisis based on Twitter Analysis," in *EGU General Assembly 2013,* Vienna, Austria. 07.-12.04.2013.